

Spotify User Behavior Dataset

About Dataset

This dataset contains information about user behavior on the music streaming platform Spotify. It includes data on listening history, user preferences, and demographic information. The dataset can be used to analyze trends in user behavior and inform recommendations for music and advertising.

The Spotify User Behavior Dataset is a comprehensive collection of anonymized data that offers valuable insights into the behavior patterns and preferences of Spotify users. This primary dataset is specifically designed for analyzing user interactions, music consumption habits, and engagement metrics within the Spotify music streaming platform. It provides researchers, data analysts, and data scientists with a rich resource to explore and understand user behavior in depth.

There are 20 column in the table:

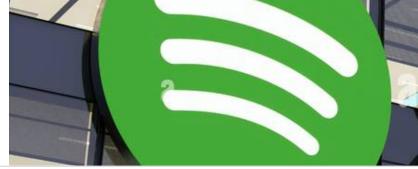
- 1) Age - Age group of user?
- 2) Gender - Gender of user?
- 3) `spotify_usage_period` - How long have you been using Spotify?
- 4) `spotify_listening_device` - Which of the following devices do you primarily use to listen to Spotify?
- 5) `subscription_plan` - Which Spotify subscription plan do you currently have?
- 6) `premium_sub_willingness` - Are you willing to take a premium subscription or willing to continue with premium subscription in future?
- 7) `preferred_premium_plan` - If premium or willing to take premium, what amount do you pay for the subscription?
- 8) `preferred_listening_content` - What do you prefer to listen more?
- 9) `fav_music_genre` - What genre(s) of music do you enjoy the most?
- 10) `music_time_slot` - What is your favorite time slot to listen to music?
- 11) `music_Influencial_mood` - When it comes to listening to music, which of the following moods or situations most strongly influences your choice of music?
- 12) `music_listen_frequency` - When do you listen to music more often?
- 13) `music_expl_method` - How do you discover new music on Spotify?
- 14) `music_recc_rating` - How do you rate the spotify music recommendations?
- 15) `pod_listen_frequency` - How often do you listen to Podcast?
- 16) `fav_pod_genre` - What genre(s) of Podcast do you enjoy the most?
- 17) `preferred_pod_format` - What podcast format you generally prefer?
- 18) `pod_host_preference` - Are you more inclined to listen to podcasts from unknown personalities, or do you prefer podcasts hosted by well-known individuals?
- 19) `preferred_pod_duration` - Do you prefer shorter podcast episodes (under 30 minutes) or longer episodes (over 30 minutes)

20) `pod_variety_satisfaction` - Are you satisfied with the variety and availability of podcasts on Spotify?

source:

Spotify User Behavior Dataset
A Comprehensive Spotify Dataset for User Analysis

<https://www.kaggle.com/datasets/meeraajayakumar/spotify-user-behavior-dataset>



Data cleansing

The data must be cleaned up before being processed and imported using extension SQL import into Azure Data Studio for analysis. If there are any empty cells in our data, the input data will be damaged, therefore tick the box to accept the Null value when setting the data type of each field during the import procedure. If the data is input successfully, there is no blank record.

Each record (row) is a data of 1 user, but the dataset does not have an id column of the user, so we add a field of `user_id` with the `ROW_NUMBER()` function with Over() clause and in this data set we have no unique field, so we create a unique field by the way combining all the columns, but in this case we just use the three first fields.

```
select ROW_NUMBER() OVER(ORDER by concat(Age, Gender, spotify_usage_period)) as user_id
      ,
      *
from dbo.spotify_user
```

Spotify_listening_device field data taxonomy

| spotify_listening_device | s |
|---|----|
| Computer or laptop, Smart speakers or voice assis | Fr |
| Computer or laptop, Smart speakers or voice assis | Fr |
| Computer or laptop, Wearable devices | Fr |

We see that there are records containing many different values, if put into Power BI, it will give many different legends according to the value of each cell, not divided by device.

We need to separate each value separated by ",".

First we split the values into multiple columns if columns, we use SQL on Azure data studio with the way the function is related to String and case : `SUBSTRING()`, `CHARINDEX()`, `LEN()`.

(we can do on excel easily)

```
With data_set as (SELECT ROW_NUMBER() OVER(ORDER by concat(Age, Gender, spotify_usage_period)) as user_id
      ,
      *
FROM dbo.spotify_user)

SELECT user_id, spotify_listening_device, spotify_listening_device_1
      , Case when spotify_listening_device_0 like '%,%' then SUBSTRING(spotify_listening_device_0,CHARINDEX(',', spotify_listening_device_0),LEN(spotify_listening_device_0)-1)
      else spotify_listening_device_0 end as spotify_listening_device_2
      , Case when spotify_listening_device_0 like '%,%' then LEFT(spotify_listening_device_0,CHARINDEX(',', spotify_listening_device_0, 1))
      else null end as spotify_listening_device_3
FROM
(SELECT user_id, spotify_listening_device
      , Case when spotify_listening_device like '%,%' then SUBSTRING(spotify_listening_device,CHARINDEX(',', spotify_listening_device, 1))
      else null end as spotify_listening_device_0
      ,Case when spotify_listening_device like '%,%' then LEFT(spotify_listening_device,CHARINDEX(',', spotify_listening_device, 1)-1)
      else spotify_listening_device end as spotify_listening_device_1
FROM data_set) sub
```

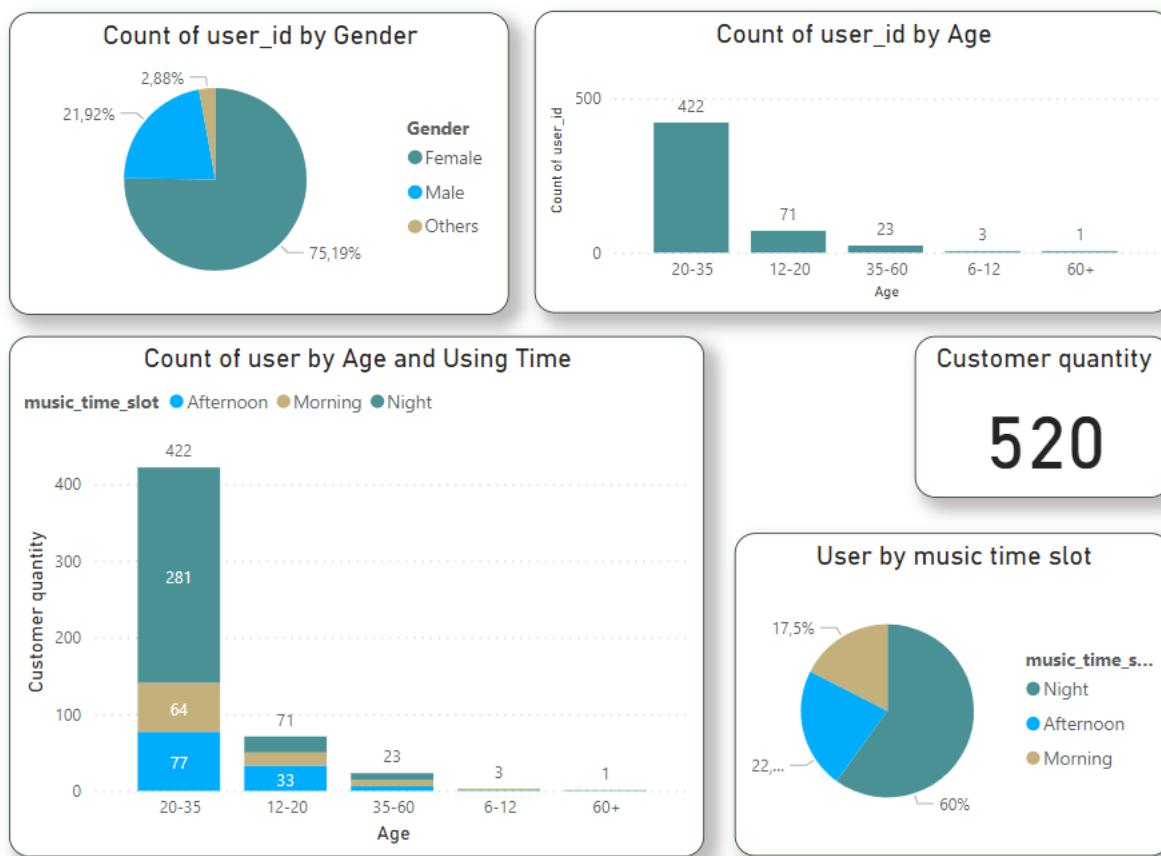
Then import into Power BI using Power Query and use the Unpivot columns function in the Transform task.

We get a sheet with each user's device information in each cell.

Same with `music_expl_method`, `music_Influencial_mood`, `music_lis_frequency` fields

Data analysis and visualization

After processing data on Azure, we bring it back to Power BI to analyze and visualize data



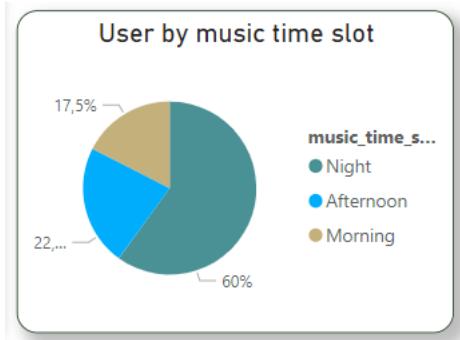
This dataset provides valuable insights into the behavior patterns and preferences of Spotify users, specifically related to their listening behavior on the platform. It includes demographic information, such as age and gender, as well as data on listening history and user preferences.

When analyzing the data based on gender, it is interesting to note that the majority of users are female, accounting for over 75% of the 520 users. In contrast, males account for over 20%, and other genders only make up a small portion of the dataset. This suggests that Spotify's user base is predominantly female, and may be useful information for companies looking to target this demographic.

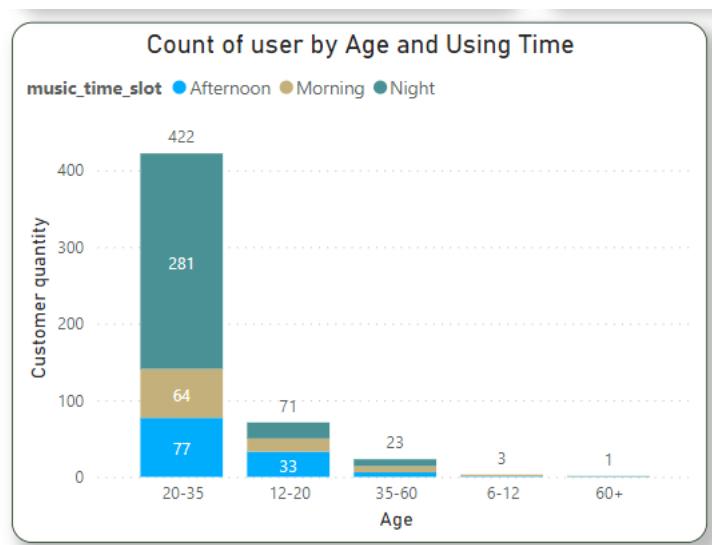
Furthermore, when looking at the age distribution, we can see that there are up to 422 users between the ages of 20 and 35, which accounts for 81.1% of the total number of users. This indicates that Spotify's user base is primarily made up of young adults who are likely to be working professionals or college students. The next largest age group is between 12 and 20, accounting for 13.7% of the users, while those aged 35 and above make up the remaining over 5%.

Overall, this data can be used to gain a deeper understanding of Spotify's user base and to make informed decisions about marketing and advertising strategies.

To better understand users' app usage habits, we've broken down three times of the day when they use the Spotify app: Evening, Afternoon, and Morning. According to our statistics, the majority of users often listen to music on Spotify in the evening, accounting for 60% of the total number



of users (equivalent to 312 users). Meanwhile, the remaining two time periods, afternoon and morning, have a fairly equal number of users, reaching 22.5% (117 users) respectively and a smaller percentage than the other two periods.

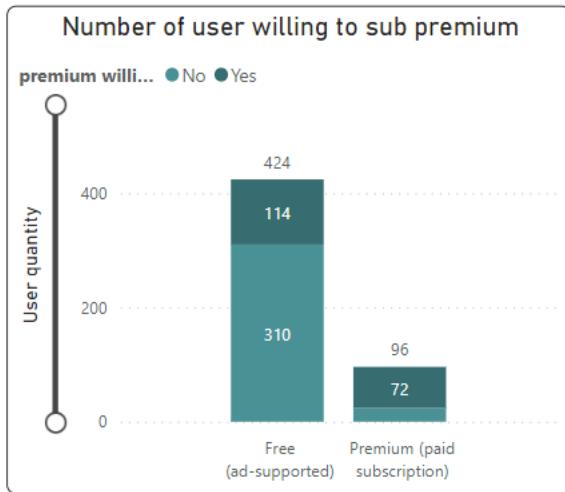


The Spotify User Behavior Dataset provides valuable insights into the behavior patterns and preferences of Spotify users, specifically related to their listening behavior on the platform. It includes demographic information, such as age and gender, as well as data on listening history and user preferences.

When analyzing the data based on age and usage time classification, we can see that a significant number of people aged between 20 and 35 use Spotify at night, possibly as a leisure time activity for them. This could be because they have finished work or their daily activities, allowing them to relax and unwind to music. For students, the usage time is quite evenly distributed, but the highest usage period is in the afternoon. This could be due to their class schedules and having more free time in the afternoon to listen to music.

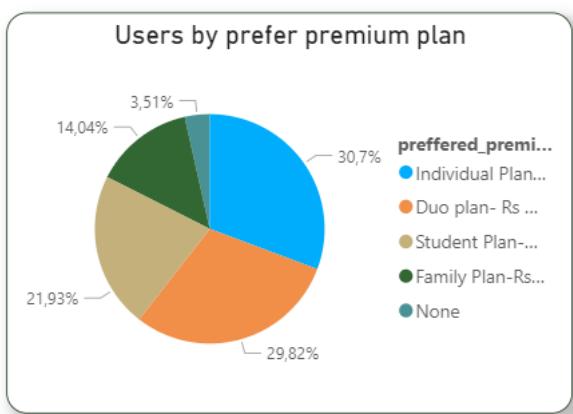
It is interesting to note that the dataset does not include information on why users prefer certain usage periods, but it can be inferred that users may have different daily routines and responsibilities that influence their listening behavior. By understanding users' listening behavior, Spotify can tailor its features and content to better meet their needs and preferences.

Overall, this information provides insight into the usage patterns of Spotify users, which can be used to inform marketing and advertising strategies, as well as product development decisions.



Of the 520 users, there are 424 Free users, accounting for more than 80% of the total users. And there are 96 users using the Premium plan and only 20%. An approximation to the **Pareto Principle (also known simply as the 80/20 Rule)**. Here 20% of users bring 80% of total revenue, can't say to bring 100% because customers using free all have ad-support and the revenue from learning is from advertising fees from other organizations, so maybe they together bring Spotify 20% of the revenue.

For free users, up to 114 people intend to register to use the Premium package



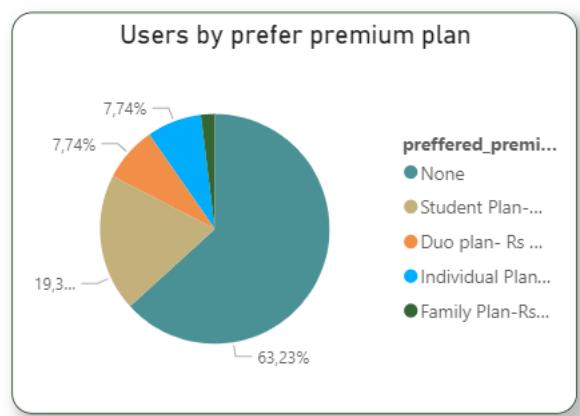
The Spotify User Behavior Dataset provides valuable information on the behavior patterns and preferences of Spotify users. Among the many insights gleaned from this dataset is the fact that a significant number of users do not intend to pay for the platform. Specifically, there are 310 users who have no intention of subscribing to any plan at all.

For these users, the majority have not yet considered or are not interested in upgrading to a paid subscription. Among those who have considered upgrading, the Student plan, which is the cheapest plan available, has the second-most number of users, with around 60. Meanwhile, the Family plan has the lowest number of users.

It's worth noting that the Student plan is a popular choice among those who want to upgrade their subscription while still

Most users of the free plan who want to upgrade to premium will choose the Individual Plan, Duo Plan, or Student Plan packages since they are significantly less expensive than the others.

Additionally, 4 actual customers who are prepared to pay for Spotify have the choice of using none; perhaps they are unsure of the best bundle to employ.

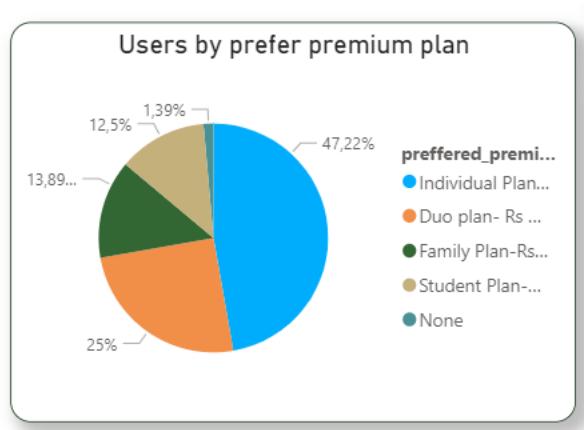


Conversely, the Family plan may be less popular among users who only need a single subscription, or who do not have a family to share the plan with.

keeping costs low. This could be because the Student plan offers many of the same features as the other plans, but at a more affordable price point.

Overall, these insights can help Spotify identify areas where they can improve their paid subscription options, as well as develop marketing strategies to encourage more users to upgrade their subscriptions. By understanding users' preferences and behaviors, Spotify can tailor their offerings to better meet the needs of their users.

For users who are using the Premium plan, up to 72 users continue to subscribe to the premium package including:



According to the Spotify User Behavior Dataset, the individual plan is likely the most popular subscription package, with 34 users planning to sign up for it. This suggests that many Spotify users prefer to have a personal plan that meets their specific music streaming needs.

Following the individual plan, the Duo plan and the Family plan are the next most popular subscription packages. This indicates that users who want to share their music streaming experience with others, such as family members or friends, may prefer to opt for these packages.

However, it's worth noting that one user is still undecided on which plan to sign up for next. This could be due to a variety of factors, such as a lack of information about the different subscription packages or uncertainty about their future music streaming needs.

Overall, these insights into Spotify's subscription plans can inform the development of targeted marketing strategies that cater to the preferences and needs of different user segments. By understanding which plans are most popular and why, Spotify can tailor its offerings to better meet the needs of its diverse user base.

Through the above analysis, we divide the total number of customers into 4 different segments:

- Paid users and continued to use that plan (72 users,)
- Paid users and do not intend to continue using (114 users)
- Free users and plan to upgrade plans (24 users)
- free user and no plan to upgrade plan (310 users)

However, we do not have any fields to divide into 4 customer files, so we need to create an additional column containing this customer classification information. You can use the available query on Azure data studio or on power PI. I will be using SQL code to generate and import the power pi, this seems to be more complicated but it helps me to practice more skills.

For SQL, it is quite easy when we use CASE WHEN to classify:

```

With data_set as (select ROW_NUMBER()
OVER(ORDER by concat(Age, Gender, spotify_usage_period))
as user_id
,
*
from dbo.spotify_user)

select user_id
, case when spotify_subscription_plan = 'Free (ad-supported)'
and premium_sub_willingness = 'Yes'
then 'Free user and will upgrade plan'
when spotify_subscription_plan = 'Free (ad-supported)'
and premium_sub_willingness = 'No'
then 'Free user and will not upgrade plan'
else 'Paid user'
end as classification
from data_set

```

```

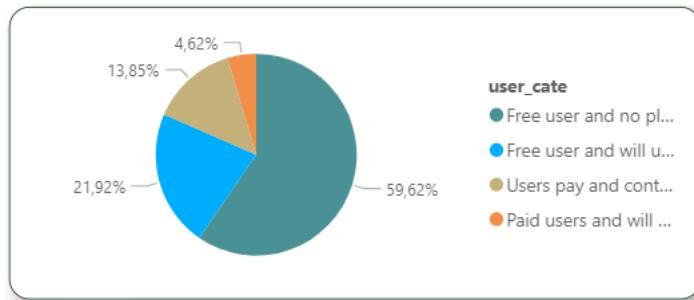
        then 'Free user and no plan upgrade'
    when spotify_subscription_plan = 'Premium (paid subscription)'
        and premium_sub_willingness = 'Yes'
        then 'Users pay and continue to subscribe'
    else 'Paid users and will not continue to subscribe'
    end as user_cate
from data_set

```

After the classification is complete, we will conduct the classification, then we proceed to export the file in excel workbook and import it into Power BI. Use Power Query to link two sheets together via user_id.

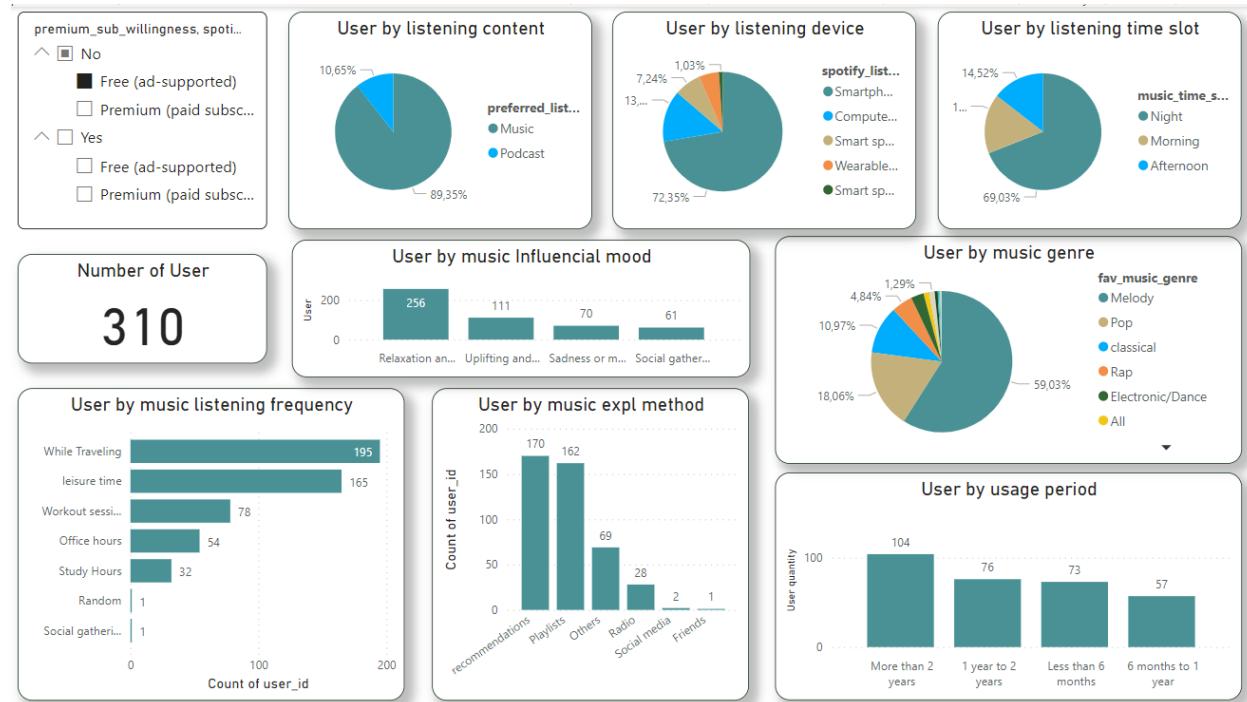
Or we can create a column always on Power BI with the following formula:

```
Column = if([spotify_subscription_plan]="Free (ad-supported)" & [premium_sub_willingness]="No", "Free user and no plan upgrade",if([spotify_
```



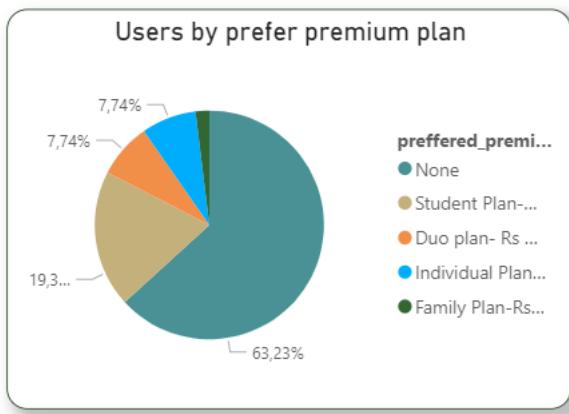
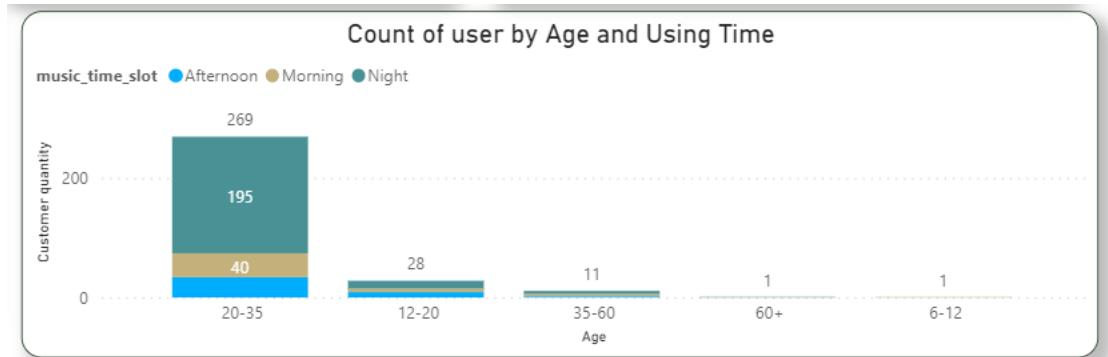
ABOUT MUSIC

Free user and no plan upgrade (Low value)



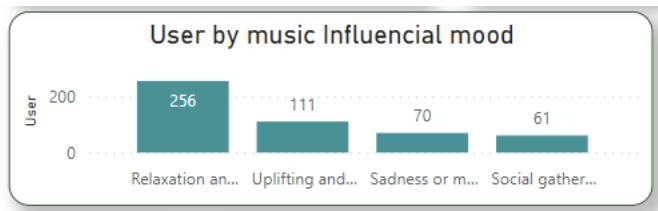
We have almost 60% (310) free users and are not ready to sign up for a premium plan. Analysis of several factors related to user behavior:

Regarding listening content, they mainly listen to music with nearly 90% of them and the remaining 10% listening to Podcast. The most used household applications are smartphones with more than 70% followed by computers or laptops. and only 4 users use on voice assistant device. Nighttime is the time when they favorite to open Spotify and listen the most.



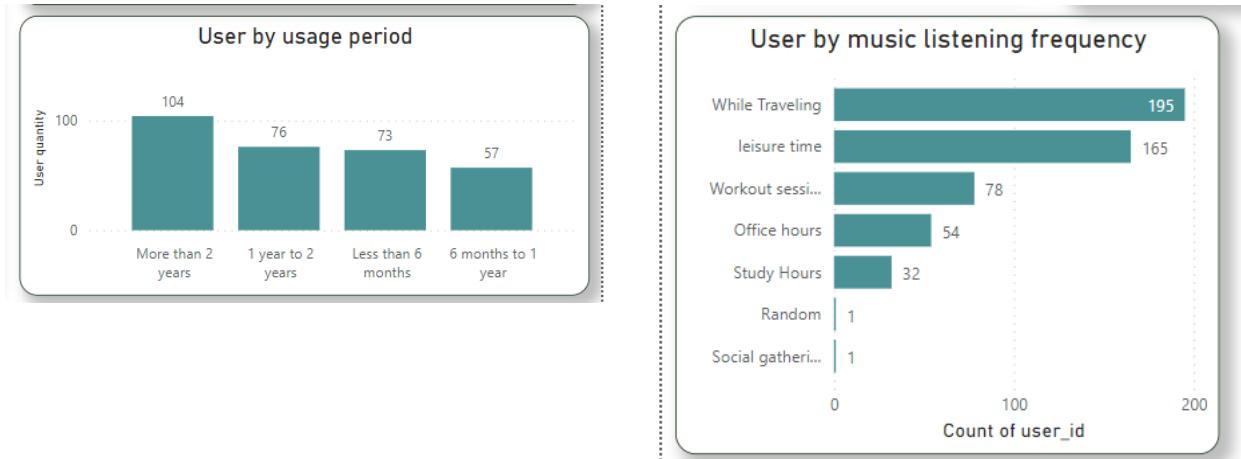
Most of them are students or workers between the ages of 20 and 35. The student package will be the one that gets the most registrations, with about 20%, as seen in the chart below. Because student identification is necessary to register for the student package, it is conceivable that the majority of this 20% group are students between the ages of 20 and 22. and the remaining choices are largely older than 22.

Nearly 60% of people chose Melody as their style of music. As far as is known, Melody has a lot of repetitious musical elements. There may or may not be gaps in the melody. Pitch, tension and projection, rhythm and form, and note associations are rhythmic combinations that make up melodies. For consumers between the ages of 20 and 35—a group that is under a lot of stress—this may be a curative potion. The music that young people adore the most is obviously POP, which comes next.



That's right, up to 256 (82.6%) users choose to listen to spotify when they feel the need to relax and feel pressured then those who need to be motivated and uplifted. 70 users listen when they feel sad or melancholy, finally 61 users listen when they are in gatherings or parties.

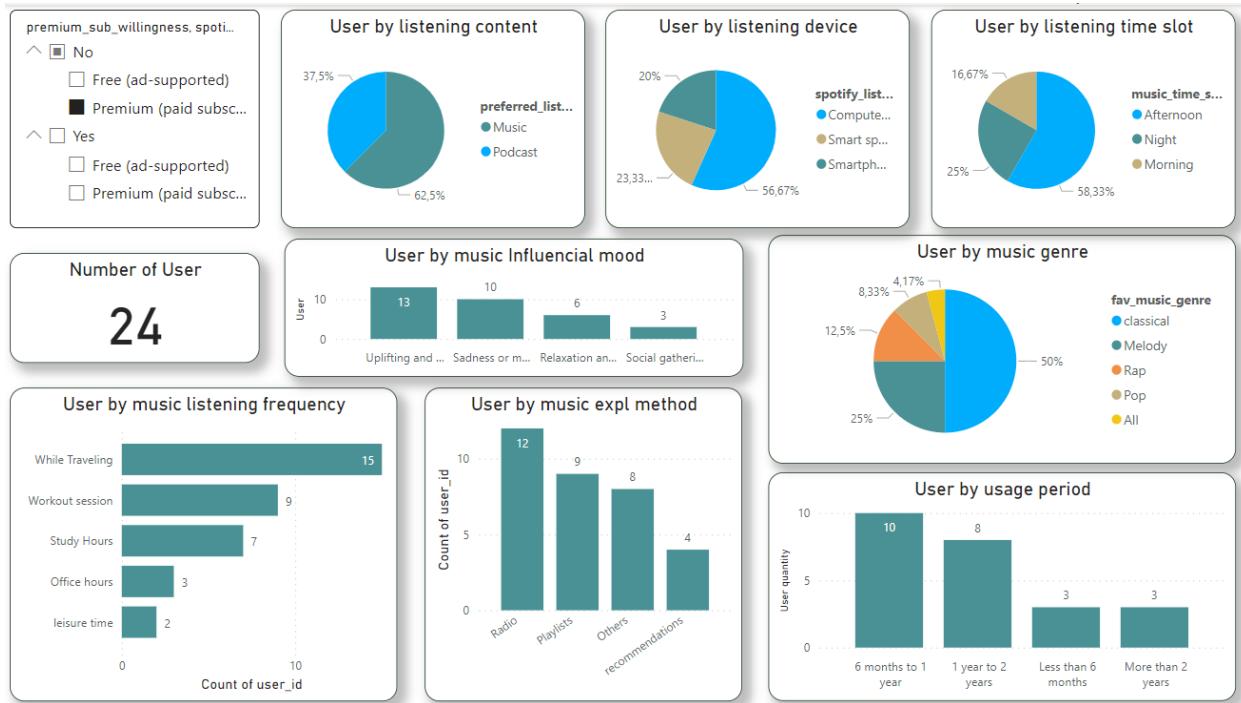
The fact that they regularly listen to Spotify while traveling with 195 options, then leisure time with 165 options, random and participate in social gatherings with only 1 choice per legends.



There are about 104 free users using Spotify for over 2 years, which is quite a long time for competition with many other platforms.

Up to 170 users trust Spotify's suggestion function to discover new music. Not much difference in second place with 162 users listening to a certain playlist to find new songs. Friends and social networks only have 1 and 2 options respectively, it can be seen that Spotify's function is quite suitable for this customer file.

Free user and will upgrade plan (potential)



Still more music than podcasts, with a slight difference compared to the above client file. This customer segment has a high usage rate on computer and laptop devices with more than 50%, but not significantly compared to other devices.

With 24 users, this data does not fully tell about this customer segment, but we should still exploit them because they are a file of potential customers. With the rate of using a laptop or computer device, we also have campanges on this device. afternoon time is also the time when they like to turn on spotify and listen.

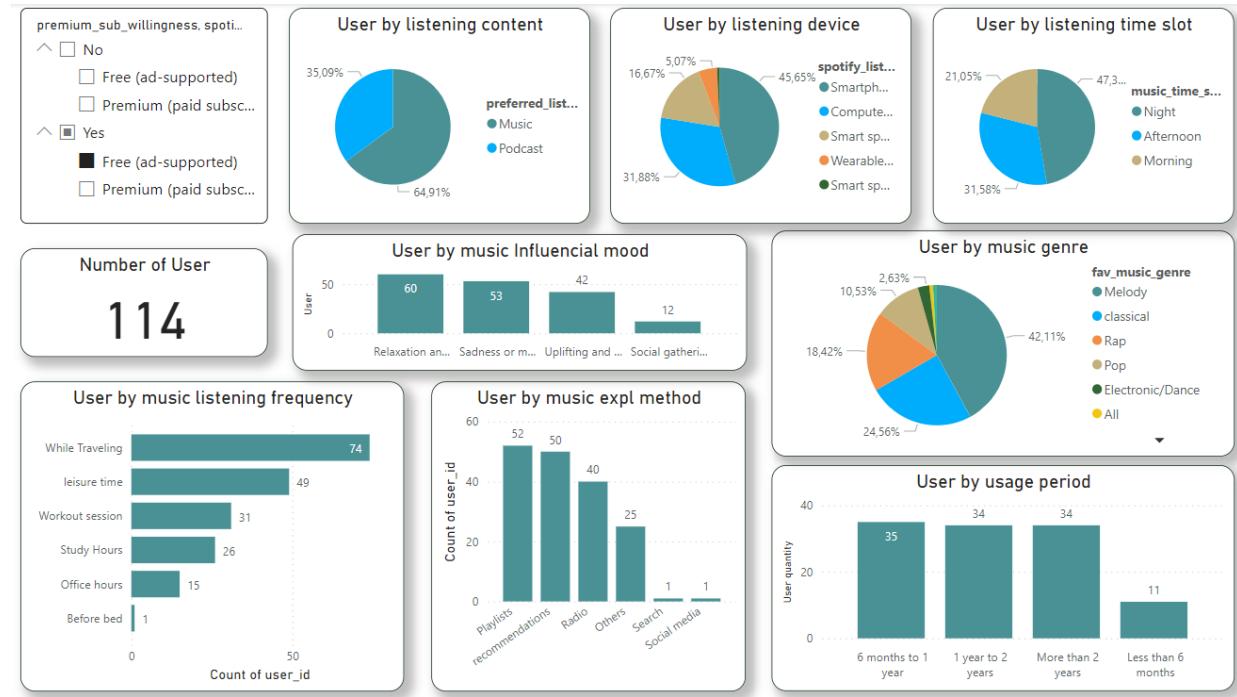
With 12 users choosing to listen to classical music, Classical Music has a calming effect , **giving you a melodious, melodious feeling**. It is recommended that you listen to classical music while working to keep your mind calm and clear. The tunes of classical

music will stimulate the production of endorphins or natural relaxation in the brain. This can have a lot of impact on this client file.

Most of the time they listen is when they are traveling with 15 choices, if these users have a choice of computer or laptop device, this is a point worth paying attention to.

Radio is the channel where they search for new songs the most with 50% of users. Next is the playlist, the song suggestion is only 4 options.

Paid users and will not continue to subscribe (About to sleep)

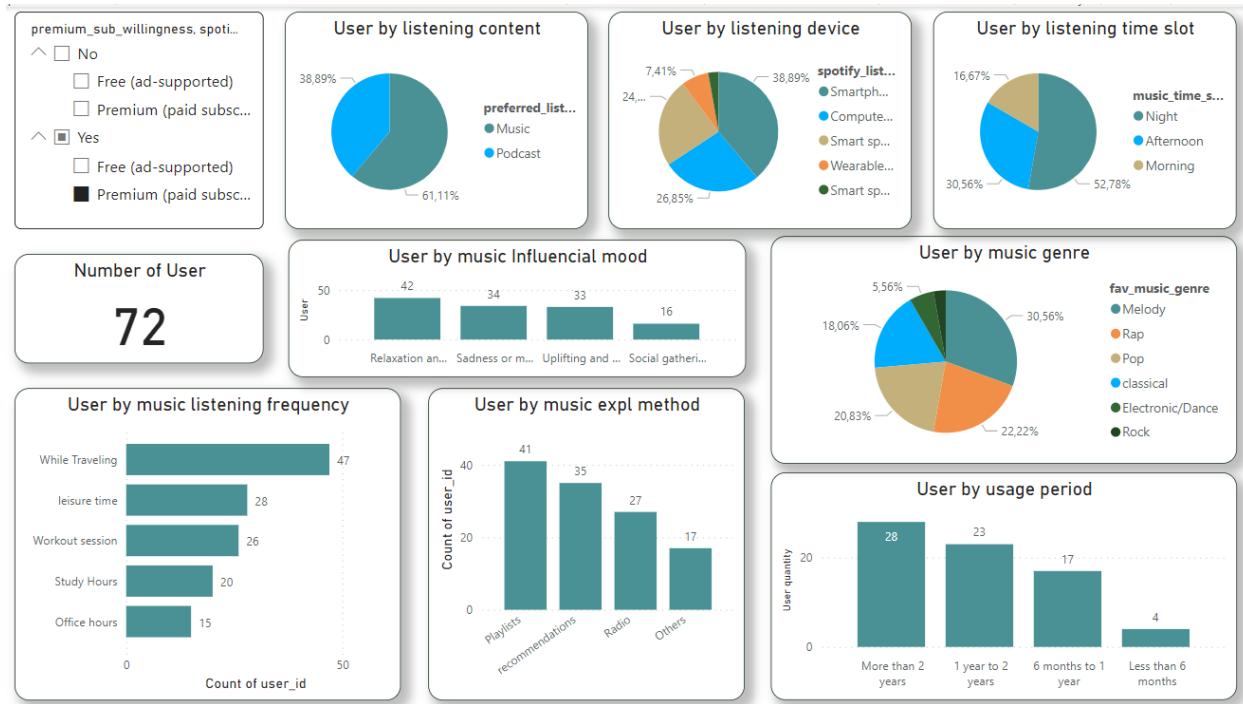


With 114 users not willing to sign up for a premium plan in the near future, this is a huge number compared to the total. In general, the behavior of this segment is no different from the Low_value file.

There is a period of time that they have used spotify which is quite long from 6 months to 2 years and more than two years accounts for a large number of users.

We should have more policies to promote this customer segment.

Users pay and continue to subscribe



With 72 loyal users of Spotify, they are still willing to sign up for premium packages in the near future, overall, nothing special compared to other segments.

ABOUT PODCAST



With a listen rate that is many times less (almost 4 times less) than music, podcasts seem to be unpopular and reach fewer spotify users, I will analyze the performance of Pods.

The percentage of users who choose the podcast channels they know and are famous for is not much different from the unknown channel owners, it seems that the channel owner does not have much influence on the channel they choose to listen to.

Up to 344 users rarely listen to podcasts on Spotify this is an extremely large number, accounting for 66.2%, followed by the number of users who never listen, the number is up to 230 users (44.2%). Only 33 users use Spotify to listen to podcasts daily.

Up to 36% of users have chosen that the duration of the podcast should be shorter, but this is not a large number to be able to confirm that the duration is a factor affecting users' listening to podcasts on Spotify. .

Up to 408 users rated the podcast listening experience on Spotify as ok ie neutral, there are 235 + 72 with the choice of satisfied and too satisfied.

With the number of 148 users who cannot listen to podcasts, it can be concluded that the majority of users do not have much access to podcasts on Spotify.